Affect of Age on Out of Pocket Health Expenditures for Breast Cancer Survivors

Georgia Institute of Technology

Carson Cole

Abstract

Medical non-adherence in breast cancer survivors can be attributed to out-of-pocket (OOP) expenditures. A cancer survivor is defined as anyone who has been diagnosed with cancer and is still living.
Some prominent variables that affect OOP spending are income, age, weight, ethnicity, etc. The relationship between OOP spending and these variables will be examined using a variety of multi linear regression analysis and their supporting tests. Discovering variables that help us understand why certain survivors pay more per year on health are valuable because we may then be more aware of when a survivor is more likely to be medically non adherent.

I. Introduction

Catastrophic health expenditures correlate with medication nonadherence and medication cost-coping strategies. WHO defines health expenditures as catastrophic when one spends more than 40% of their non-subsistence income on health services. One participates in medication non-adherence they do not seek and use medical services properly. Medication cost-coping strategies can range from borrowing money to selling medication to alleviate costs to not purchasing the medication at all.

1 in 8 American women will be diagnosed with breast cancer and the frequency has been increasing (Short et al 2010). Pisu et al (2014) estimate that OOP costs range between $300 and $1,180 per month in the first year of treatment and $500 afterwards. While these numbers do not guarantee catastrophic spending, many will face the decision of whether to seek proper medical treatment. Breast cancer patients over the age of 75 have shown to more likely to be medically non adherent (Zullig et al 2013). While there are multiple hypotheses around this, one thought is that elderly individuals do not believe the costs to be worth it given their limited time left. Supporting, Allaire et al (2017) found that OOP was higher for woman under the age of 45. They concluded that further research needs to be held to discover why this is. These two points would suggest that OOP spending is higher for younger adults, potentially because elderly patients are more likely to be medically non-adherent.

A 2013 study from Zullig et al found that breast cancer patients with higher OOP costs for their aromatase inhibitors were more likely to be non-adherent with their medication as opposed to patients with lower OOP costs. They also found that patients with higher costs were also more likely to abandon their oral chemotherapy prescriptions which increased their probability of mortality.

This information motivates research in the study of variables that affect OOP expenditures for breast cancer patients. Patients who coincide with these variables may be deemed more susceptible to medical non-adherence and could benefit from additional monitoring and briefing .

Hypothesis

Out-of-pocket spending for breast cancer patients will be the dependent variable. Age is the primary independent variable. Younger persons may put a higher value on their health and thus be more willing to pay more to maintain their health. Older persons may view these treatments as having a diminishing return on extending the years until they pass. A person, say 85, may not think it worth to pay $16,910 in the initial year of treatment (average OOP expenditure during the first year according to Short et al

2010). Instead they may believe their spouse, children, or organization could reap better benefits from this money. Hypothetically, OOP expenditures and age will have an inverse relationship. A higher age will result in lower OOP spending.

This hypothesis will be used to test each variable

$$H0 : \beta k = 0$$

$$H1 : \beta k \neq 0$$

Where Bk represents any of the independent variables. A two tailed t-test will be used to test the significance.

II. Literature Review

A study to estimate national OOP medical expenditures for cancer survivors whose age was less than 65 was conducted by Short et al in 2010. This study only analyzed those under the age of 65 because the majority of those above that age are under Medicare and should be viewed in a separate analysis. Data from 2001 to 2007 was gathered from the Medical Expenditure Panel Survey (MEPS) and the National Health Interview Survey (NHIS). Propensity score matching was used to estimate the relationship between cancer and OOP expenditures. Probit models were developed to estimate the probability of exceeding various expenditure thresholds. Short et al found that the average annual OOP expenditures for cancer survivors was $16,910 for those who had been diagnosed that year and $7992 for those diagnosed in previous years. The average for all survivors was approximately $9300. The authors found this study to be of importance for a few reasons. Cancer diagnoses have increased from 6 to 12 million in the past 20 years. These individuals who are given diagnosis have higher life expectancies due to advancements in treatment, and they are bringing into question the long term effects of these treatments and whether or not they are worth it. Cancer has been treated with increased combinations of higher doses across a wide range of medical procedures such as radiation, hormone, and chemo therapies. Potential long term consequences of treatment have shown to be “impaired physical and organ function, changes in appearance, sexual dysfunction, incontinence, lymphedema, hormone imbalances, cognitive difficulties, and fatigue.”

Kaisaeng et al (2014) conducted a cross-sectional random sample study of 5% of Medicare beneficiaries who were prescribed imatinib, erlotinib, anastrozole, letrozole, or thalidomide to treat their oral cancer in 2008. Discontinuation or delay of medication was measured against OOP expenditures using logistic regression. The patients studied in the sample did not receive subsidies for their medication and were aged 65 or older. OOP costs for medication ranged from $2.96-$37.47 per day depending on the medication. The study found that an increase in OOP costs resulted in an increase chance that the individual discontinued or delayed their medication. Variables included in their analysis included age, sex, Caucasian (binary), number of comorbidities, number of non cancer drugs, beneficiaries with enhanced alternative benefit, and cost of non cancer drugs. Medication delay was defined as going 30 days or more “between the date that the patient’s supply of the medication should have expired and the date that the patient obtained the next refill” Discontinuation was defined as a patient being without medication for more than 30 days. Overall, this paper finds that increases in OOP spending increases the probability that an oral cancer patient will be medically non-adherent with their medications.

Bernard et al conducted a study in 2011 using NHIS data from 2001-2009. They compared the ratio of OOP expenditures and income for adults between the ages of 18 and 64 who received treatment for cancer. They then looked at the proportion of these patients who lived in lived in high health burden households. They defined a high health burden as spending more than 20% of income on health related services. They found that 13.4% of these cancer patients had high health burdens while 9.7% of other chronic illness patients did, and 4.4% of those without chronic illnesses experienced this. Among those under the age of 65 without cancer, they found the highest correlation between high health burdens and age “55-64, non-Hispanic black, never married or widowed, one child or no children, unemployed, lower income, lower education level, living in nonmetropolitan statistical areas, and having other chronic conditions.” They then categorized the patients by age in groups of 18-39, 40-54, and 55-64. They found that older age groups were more likely to experience high health burdens. They also categorized based on whether they were insured with a private group, private nongroup, public, or uninsured. They found private groups had the lowest chance of being in a high health burden household while private non-groups had the highest.

To the best of my knowledge, no other paper has used 2019 NHIS data to examine the effect of age on OOP health spending across all age groups instead of being split at the age of 65. I will conduct a chow test to see if an MLR may be able to pool both age groups, those under and over 65. My hope is that this contributes to the understanding of how age affects OOP spending. If it does, then older or younger breast cancer patients with more or less time passed since their diagnosis may need to be given additional monitoring and briefing to ensure that the costs associated with their treatment does not lead to medical non-adherence.

III. Data

Variables are included in this analysis for the following reasons. Employment and family income were found to be significant factors affecting OOP spending according to Rashidul et al (2017). Insurance coverage was found to impact OOP spending in Bernard et al (2014). Hispanic ethnicity is included because Bustamante et al (2011) found those of Hispanic ethnicity had lower OOP. Obesity leads to a number of health risks and thus weight is included. Medicare coverage is included because there is a drop off in OOP at the age of 65 and this is likely due to Medicare coverage. GTE65 is also included to conduct the chow test. GTE65 and Medicare could likely be used interchangeably, but both will be used in separate analysis for specific purposes.

Table.1 Variable Discriptions

Variable Name	**Description**	**Type of Values**
Ln(FAMINCTC_A)	Natural log of family income	Integer ($)
AGEP_A	Age	Integer
WEIGHTLBTC_A	Weight	Integer
HISP_A	Hispanic ethnicity	1 or 0
NOTCOV_A	Has health insurance	1 or 0
EMPWRKLSWK_A	Worked last week	1 or 0
HICOSTR1_A	Out-of-pocket health expenditures	Integer ($)
medicare	Has Medicare insurance	1 or 0
GTE65	Age greater than are equal to 65	1 or 0

Source

The source of the data is the 2019 CDC National Health Interview Survey. The original sample size was 31,997. Responses coded as “Refused to answer”, “Does not know”, or similar have been dropped from the data. Observations were dropped for those who had not had breast cancer. The sample size used in this study is 272. The variable ln(FAMINCTC_A) was created from FAMINCTC_A. GTE65 is a binary variable created from AGEP_A where survivors were coded as 1 for being 65 or older and 0 for younger.

Table.2 Variable summary

Variable	Obs	Mean	STD	Min	Max
HICOSTR1_A	272	3,709.48	4,336.41	1	40,000
FAMINCTC_A	272	62,432.98	51,492.52	3,000	220,000
ln(FAMINCTC_A)	272	10.72	0.83	8.01	12.301
AGEP_A	272	69.48	11.60	23	85
WEIGHTLBTC_A	272	163.59	33.17	100	270
HISP_A	272	0.07	0.26	0	1
NOTCOV_A	272	0.02	0.13	0	1
EMPWRKLSWK_A	272	0.71	0.45	0	1
medicare	272	0.71	0.45	0	1

CLM Assumptions

Here, the six classic linear model assumptions will be examined to determine whether we may carry on to forming valid regressions.

MLR.1: Model is linear in its parameters. We might be able to predict OOPSpending given these variables which are linear parameters. Each parameter is linear because the degree of its polynomial is 1. The u represents the error term.

$$\text{HICOSTR1\_A} = \beta_1 ln(FAMINCTC\_A) + \beta_2(AGEP\_A) + \beta_3(WEIGHTLBTC\_A) + \beta_4 (HISP\_A) + \beta_5(EMPWRKLSWK\_A) + \beta_6(medicare) + u$$

MLR.2: Data is gathered using random sampling. The data is gathered using geographically clustered sampling, so it is not completely random and observations are not independent of each other. Although, the sample has been obtained in a manner that the CDC believes to be representative of the nation. Also, the mean age of this sample is 69, far above the average US citizen age of 38. Even still, this analysis may still produce a useful conclusion.

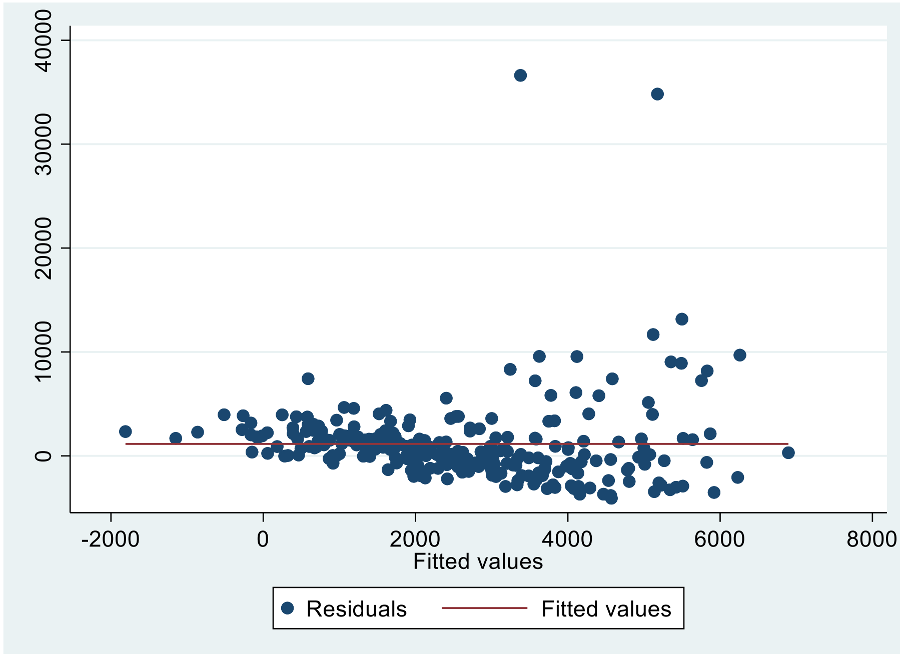
Table.2 Regions from which participants lived

Region	Northeast	Midwest	South	West
Percent	16.91	22.20	36.49	24.40

MLR.3: No perfect collinearity. During the regression, NOTCOV_A (whether or not someone had health insurance) was found to have an issue with collinearity due to almost all participants having health insurance. In order to maintain the integrity of the regressions, NOTCOV_A has been committed. After reviewing the correlation table, STATA.2, we can see that none of the other variables are linear combinations of each other, nor do they demonstrate prefect collinearity. A high correlation was found between AGEP_A and medicare due to medicare only being available to those who are 65 and older. This will provide an issue with multi collinearity when performing the Chow test. Thus, the variable medicare will be removed for the Chow test. We can also see that the variance inflation factor (VIF) is relatively low for all of the variables with the highest being 3.01 for AGEP_A. AGEP_A is thought to have multi-collinearity with the presence of medicare and lack of employment; a relationship which will be examined in the regressions.

MLR.4: **Zero Conditional Mean.** There is a chance of misspecification in the MLR. A variable such as height has been shown to have effects on health, but it is not being used in this analysis due to its low impact. If height were to be included it would have a correlation with weight such that $\delta 1$!= 0 Thus, we cannot guarantee that E(U|**x**) != 0. In image.1 , we can see that the average residual hovers slightly above 0 and does not change as the fitted values increase. Thus, the MLR.4 assumption is relatively strong even though we are not accounting for all variables that could be potentially affecting HICOSTR1_A.

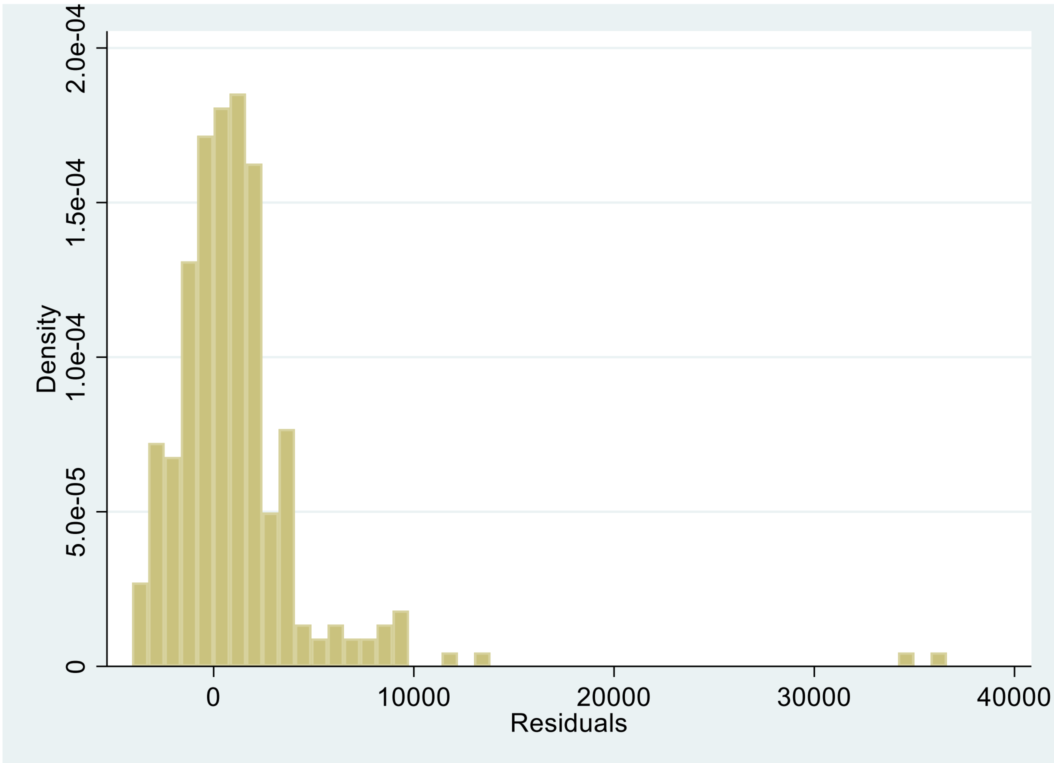
Figure.1 Residuals



MLR.5: Homoskedasticity. There could be a term in U that has a change in variation depending on the value of **x,** so we cannot be certain that MLR.5 is true. We can see that the residuals grow larger as the fitted values increase. This is a signal that heteroskedasticity may be present in the data. The breusch-pagan test in STATA.3 confirmed heteroskedasticity with a p-value < 1%. Thus, MLR.5 is violated and we cannot assume that the residuals do not vary with **x.**

MLR.6: Normality. Because there are many unobserved variables in the U term affecting Y, we could assume that U has a normal distribution due to the central limit theorem. We prove this further by plotting a histogram in figure.2 that shows that the residuals have an approximately normal distribution with a slight positive skew. This implies that y is also following an approximately normal distribution.

Figure.2 Histogram of Residuals



III. Results

After reviewing the CLM's, we can conclude that not all of them are perfectly satisfied. Although, constructing regressions and performing tests on this data may still prove useful.

Simple Linear Regression Model

HICOSTR1_A = 9165.07 - 79.80(AGEP_A) + u

From the model in STATA.4, for every year a recipient aged, on average, their OOP expenditures would decrease by $79.80. A 25 year old would be expected to spend $7170, while someone who was 85 would be expected to spend $2382. The coefficient of the AGEP_A variable is significant at <1%. This aligns with the prediction that older persons are likely to spend less on OOP health and the literature cited above. This regression produces an R-square value of 0.0484 which motivates a multi linear regression in hopes to account for more of the variance in OOP and to find a less biased coefficient of age in a regression.

Multi Linear Regression Model 1

$$HICOSTR1\_A = 5960.25 - 12.75(AGEP\_A) + 1144.90\ln(FAMINCTC\_A) - 16.72(WEIGHTLBTC\_A) + 2041.12(HISP\_A) - 891.60(EMPWRKLSWK\_A) - 1172.90(medicare) + u$$

From STATA.5: Similarly to the SLR, we see that an increase in age decreases OOP spending but by significantly less. Furthermore, age is now nowhere close to being significant. For every 1% increase in family income, OOP increased by about $11.45 with significance at 1%. Weight has an inverse relationship where for every pound added to weight, they spend $16.71 less with significance at 5%. Not having Hispanic ethnicity increased OOP by $2041.12, significant at 10%. Having worked the previous week was not significant and neither was having medicare. This produces an R-squared value of .1387, marginally better than our SLR model. This indicates that there are other variables other than the ones in this MLR that could better account for the variability in a person's OOP.

Multi Linear Regression Model 2

$$HICOSTR1\_A = -4686.11 - 71.91(AGEP\_A) + 1287.66\ln(FAMINCTC\_A) - 17.17(WEIGHTLBTC\_A) + 2055.40(HISP\_A) + u$$

From STATA.6: After dropping employment and medicare, the insignificant variables, we see similar coefficients and significance levels for all variables except a persons age. Age jumps to a similar significance and coefficient as we saw in the SLR. As individuals age they are less likely to work and more likely to have medicare. This leads to the belief that dropping the medicare and employment variables between MLR1 and MLR2 are likely what caused age to become significant with a large coefficient due to multi-collinearity between these three variables.

Multi Linear Regression Model 3

$$HICOSTR1\_A = 5232.7 + 2.03(AGEP\_A) - 956.89(EMPWRKLSWK\_A) - 1752.91(medicare) + u$$

From STATA.7: We now wish to look at the regression with age along with the two variables it is believed to have multicollinearity with, employment and medicare. We now see that age becomes entirely insignifigant. Employment did not become signifigant while medicare became signifigant at 5%. In this model, if one has medicare they are expected to spend $1752.91 less on OOP. We could conclude that of the three, medicare status is a better predictor of a person's OOP health costs.

Multi Linear Regression Model 4

$$HICOSTR1\_A = 5875.60. - 12.31(AGEP\_A) - 2131.87(medicare) + u$$

From STATA.8: After removing the insignifigant variable in the last model, employment, age still remains insignifigant. On the other hand, medicare now reaches signifigance of 1% with an increased coefficient. In this model, medicare patients are expected to spend $2131.87 less on OOP health expenditures. We still conclude that medicare status is a better predictor for OOP than age.

Table.3 Regression Table, Coefficients, and Signifigance

Independent Variables	SLR	MLR1	MLR2	MLR3	MLR4
ln(FAMINCTC_A)		1144.90*** (380.18)	1287.66*** (376.68)		
AGEP_A	(-79.80)*** (21.53)	-12.74 (35.89)	-71.91*** (23.34)	2.03 (35.71)	-12.31 (33.98)
WEIGHTLBTC_A		-16.71** (7.62)	-17.17** (7.61)		
HISP_A		-2041.12** (1104.66)	-2055.40** (1108.54)		
EMPWRKLSWK_A		-891.60 (725.43)		-956.89 (740.51)	
medicare		-1172.90 (870.19)		-1752.91** (884.91)	-2131.87*** (835.95)
Intercept	9165.07	-5960.25	-4686.105	5232.7	5875.60
R^2	0.0484	0.1387	0.1227	.0766	0.0709
Adj. R^2	0.0449	0.1192	0.1095	.0663	0.0639

*10%, **5%, ***1% signifigance

Table.4 MLR1 Variable Output Details

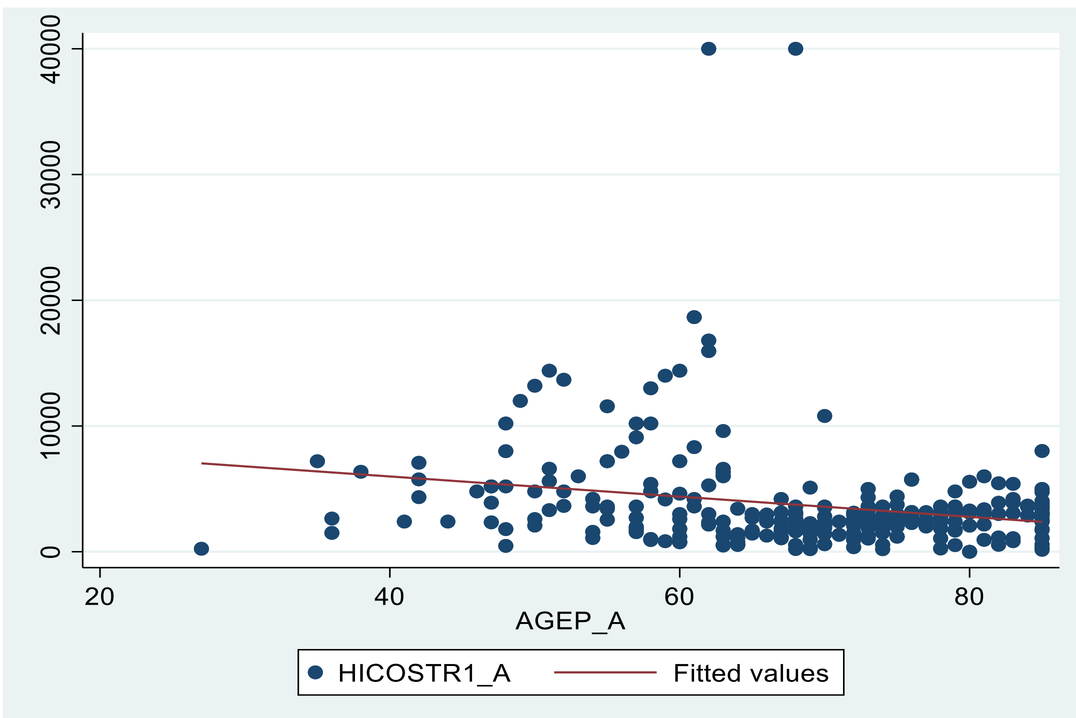
Variable	Coefficient	t-value	p-value	95% CI
FAMINCTC_A	1144.90	3.01	0.003	(396.34, 1893.47)
AGEP_A	-12.74	-0.36	0.723	(-83.42, 57.92)

WEIGHTLBTC_A	-16.71	7.62	-2.20	(-31.74, -1.73)
HISP_A	-2041.12	-1.85	0.07	(-4216.15, 133.92)
EMPWRKLSWK_A	-891.60	-1.23	0.220	(-2319.95, 536.74)
medicare	-1172.90	-1.35	0.179	(-2886.27, 540.48)

IV. Extensions

Chow Test

Figure.3 Visualization of OOP Regressed on Age



In Figure.3, we notice a decline in the amount of OOP and its variability once persons cross the age of 65. It also seems as though with a higher ages comes higher OOP with a drop at 65 and then a small increase in OOP as age continues to progress. This is contrary to the belief that with higher age comes lower OOP. This is believed to be attributed to medicare becoming available at this age. Two regressions will be ran and then used in a chow test to see if they can be pooled together. The first regression will be the MLR1 of participants aged 65 and older, while the second regression will be MLR1 of those under 65.

The pooled regression will be MLR1. We are interested in if the split regressions can be represented or 'pooled' by the pooled regression. All three of these MLRs will have a common deviation from MLR1. Since Age and medicare have such a high correlation (.7272), the medicare variable will be dropped from each regression.

*Note: the Chow test is only valid under the assumption of homoscedasticity. Even though we found heteroskedasticity in the brusch-pagan test, we will still run the chow test and interpret the results as if homoskedasticity was present.

H0 : We cannot pool all age groups into a regression

H1 : All age groups can be pooled into a regression

MLR.Chow.Pool

$$HICOSTR1\_A = -2840.36 - 41.19(AGEP\_A) + 1190.97\ln(FAMINCTC\_A) - 17.80(WEIGHTLBTC\_A) + 2119.54(HISP\_A) - 1208.54(EMPWRKLSWK\_A) + u$$

From STATA.9: After dropping medicare from MLR1, the only notable change is that employment becomes signifigant at 10% and indicates that if someone worked in the past week, they will spend 1208.54 less on OOP on average.

MLR.Chow. >= 65

$$HICOSTR1\_A = -2961.44 - 16.24(AGEP\_A) + 577.12\ln(FAMINCTC\_A) - 6.25(WEIGHTLBTC\_A) + 676.13(HISP\_A) - 853.79(EMPWRKLSWK\_A) + u$$

From STATA.10: We see all variables lose signifigance in this model.

MLR.Chow. < 65

$$HICOSTR1\_A = -20029.53 - 43.67(AGEP\_A) + 2511.77\ln(FAMINCTC\_A) - 30.13(WEIGHTLBTC\_A) + 2434.29(HISP\_A) - 518.83(EMPWRKLSWK\_A) + u$$

From STATA.11: Both ln(income) and weight remain at their signifigance levels of 1% and 5% respectively, but now their coefficients have doubled. For every 1% increase in income a person is expected to spend $25 less on OOP and for every pound gained, a person is expected to spend $30 less on OOP. It should also be noted that the intercept has shifted to about -20,000, 5 times its value in MLR1.

F-Value

The Chow value was then computed using the following equation.

$$Chow = \frac{[SSR_{pool} - (SSR_{over65} + SSR_{under65})}{SSR_{over65} + SSR_{under65}} \cdot \frac{n-2(k+1)}{(k+1)} = 2.66$$

Using degrees of freedom 5 and 266 for the numerator and denominator respectively, we find this result to be significant at 5%. Thus, we reject the null hypothesis with 95% confidence and conclude that the pooled regression is sufficiently representative for those under and greater than or equal to the age of 65.

Table.5 Chow Test **Coefficients, and Significance**

Independent Variables	Pool	Age >= 65	Age < 65
ln(FAMINCTC_A)	1190.97*** (379.23)	577.12 (353.71)	2511.77*** (911.69)
AGEP_A	-41.19 (29.08)	16.24 (41.91)	43.67 (75.36)
WEIGHTLBTC_A	-17.80** (7.58)	-6.25 (7.93)	-30.13** (-14.89)
HISP_A	-2119.54* (1104.82)	-676.13 (1325.41)	-2434.29 (1913.05)
EMPWRKLSWK_A	-1208.54* (687.32)	-853.79 (686.04)	-518.83 (1547.81)
Intercept	-2840.36	-2961.44	-20029.53
R^2	0.1327	0.0339	0.1478
Adj. R^2	0.1164	0.0057	0.0999

*10%, **5%, ***1% signifigance

V. Conclusions

The chow test had two purposes. Previous studies have conducted analysis either over or under the age of 65 because they believe that the two age groups cannot be represented by one MLR. Thus one purpose whether or this was true. The reasoning behind the splitting is that there is a belief that there will be a signifigant difference in OOP health expenditures based on whether you are over or under the age of 65, attributed to medicare availability. Thus the other purpose was to see if the coefficients for age in the split groups differed signifigantly from the pooled MLR. Although the coefficients for age in the pool were negative and positive for both splits, the age coefficient in the pooled was significant while neither split age groups yielded remotely signifigatn coefficients for age. Therefore, we can reject the null hypothesis (as thought in Bernard et al 2011) that have the thought that both age groups cannot be pooled. Although, this rejection is expected to have an increased type 1 error due to the fact that not all MLRs could be assumed. Specifically, MLR.5 could not be assumed due to the presence of heteroskedasticity as determined by the brusch-pagan test. Thus, this rejection should not be taken on concrete grounds.

Now that we have established the relationship between age, OOP, and age groups split at 65, we can discuss our original hypothesis as to whether age is a signifigantly signifigant factor in a breast cancer survivor's annual out-of-pocket health expenditure. According to the pooled model used in the chow test, age is a signifigant factor. Thus, we reject the null hypothesis that age is not a signifigant factor in a survivors OOP.

One variable that would have been interesting to see is if one became medically non-adherent. At first thought, one would believe the presence of this variable would drastically drop OOP health expenditures, but I believe the relationship would be more complex. Being medically non adherent for one medical procedure does not mean you are medically nonadherent for all. You may be okay with pills but don't want to go through radiation. Another such as whether one became medically non would have been useful to this analysis but we were restricted by the available variables in this data set.

References

Pisu, Maria et al. "The out of pocket cost of breast cancer survivors: a review." *Journal of cancer survivorship : research and practice* vol. 4,3 (2010): 202-9. doi:10.1007/s11764-010-0125-y

Sedjo, R.L., Devine, S. Predictors of non-adherence to aromatase inhibitors among commercially insured women with breast cancer. *Breast Cancer Res Treat* **125,** 191–200 (2011). https://doi.org/10.1007/s10549-010-0952-6

Mahumud, Rashidul Alam et al. "Distribution and Determinants of Out-of-pocket Healthcare Expenditures in Bangladesh." *Journal of preventive medicine and public health = Yebang Uihakhoe chi* vol. 50,2 (2017): 91-99. doi:10.3961/jpmph.16.089

Bernard, Didem S M et al. "National estimates of out-of-pocket health care expenditure burdens among nonelderly adults with cancer: 2001 to 2008." *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* vol. 29,20 (2011): 2821-6. doi:10.1200/JCO.2010.33.0522

Bustamante, A. V., & Chen, J. (2011, June 3). *Health Expenditure Dynamics and years of U.S. residence: Analyzing spending disparities among Latinos by citizenship/nativity status.* Wiley Online Library. Retrieved October 20, 2021, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6773.2011.01278.x?casa_token=2MnLmHdR7kEAAAAA%3A7OL9H8iSiyrh2BZwQCUuuTp0GEVYaYw2jLv6KplqH1dEsWLWMKJYUzGm8yCAyRsZnYtkVw_xND_RQVJj.

Senturk, B., Isik, A. D. & Gokmen, E. C. (2019). Determinants of out of pocket healthcare expenditures: The case ofMugla province in Turkey.

Global Journal of Business, Economics and Management: Current Issues. 9 (2),76-83.https://doi.org/10.18844/gjbem.v9i2.4225

Appendix

STATA.1 Variance Inflation Factor

Variable	VIF	1/VIF
AGEP_A	3.01	0.332041
medicare	2.93	0.341802
EMPWRKLSWK_A	2.08	0.481862
lnFAMINCTC_A	1.19	0.839176
WEIGHTLBTC_A	1.11	0.903108
HISP_A	1.04	0.957742
Mean VIF	1.89	

STATA.2 Corelation Coefficients

	HICO~1_A	lnFAMI~A	WEIGHT~A	HISP_A	AGEP_A	EMPW~K_A	medicare
HICOSTR1_A	1.0000						
lnFAMINCTC_A	0.2821	1.0000					
WEIGHTLBTC_A	-0.0969	-0.0625	1.0000				
HISP_A	0.0599	-0.0375	0.0538	1.0000			
AGEP_A	-0.2200	-0.3375	-0.2304	0.1681	1.0000		
EMPWRKLSWK_A	-0.2374	-0.3232	-0.1703	0.1311	0.6703	1.0000	
medicare	-0.2653	-0.3498	-0.1172	0.1102	0.7787	0.6763	1.0000

STATA.3 Heteroskedasticity Test

```
. hettest

Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of HICOSTR1_A

H0: Constant variance

    chi2(1) = 126.09
Prob > chi2 = 0.0000
```

STATA.4

```
. regress HICOSTR1_A AGEP_A

      Source |       SS           df       MS      Number of obs   =       272
-------------+----------------------------------   F(1, 270)       =     13.73
       Model |   246577182         1   246577182   Prob > F        =    0.0003
    Residual |  4.8494e+09       270  17960887.6   R-squared       =    0.0484
-------------+----------------------------------   Adj R-squared   =    0.0449
       Total |  5.0960e+09       271  18804490.1   Root MSE        =      4238

------------------------------------------------------------------------------
  HICOSTR1_A | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      AGEP_A |  -79.79775   21.53665    -3.71   0.000    -122.1989   -37.39663
       _cons |   9165.073   1494.665     6.13   0.000     6222.393    12107.75
------------------------------------------------------------------------------
```

STATA.5

```
      Source |       SS           df       MS      Number of obs   =       272
-------------+----------------------------------   F(6, 265)       =      7.11
       Model |   706577521         6   117762920   Prob > F        =    0.0000
    Residual |  4.3894e+09       265  16563921.9   R-squared       =    0.1387
-------------+----------------------------------   Adj R-squared   =    0.1192
       Total |  5.0960e+09       271  18804490.1   Root MSE        =    4069.9

------------------------------------------------------------------------------
  HICOSTR1_A | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
lnFAMINCTC_A |   1144.904   380.1844     3.01   0.003     396.3371     1893.47
WEIGHTLBTC_A |  -16.71998   7.615388    -2.20   0.029    -31.71434   -1.725613
      HISP_A |  -2041.117   1104.662    -1.85   0.066    -4216.149    133.9152
      AGEP_A |  -12.74768   35.89219    -0.36   0.723    -83.41784    57.92249
EMPWRKLSWK_A |  -891.6026   725.4337    -1.23   0.220     -2319.95    536.7447
    medicare |  -1172.896   870.1948    -1.35   0.179    -2886.271    540.4795
       _cons |  -3919.131   5390.166    -0.73   0.468    -14532.13     6693.87
------------------------------------------------------------------------------
```

STATA.6

```
      Source |       SS           df       MS      Number of obs   =       272
-------------+----------------------------------   F(4, 267)       =      9.33
       Model |   625117138         4   156279284   Prob > F        =    0.0000
    Residual |  4.4709e+09       267  16744942.6   R-squared       =    0.1227
-------------+----------------------------------   Adj R-squared   =    0.1095
       Total |  5.0960e+09       271  18804490.1   Root MSE        =    4092.1

------------------------------------------------------------------------------
  HICOSTR1_A | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
lnFAMINCTC_A |   1287.664   376.6844     3.42   0.001     546.0142    2029.314
WEIGHTLBTC_A |  -17.17192   7.605725    -2.26   0.025    -32.14675   -2.197097
      HISP_A |  -2055.401   1108.536    -1.85   0.065    -4237.984    127.1819
      AGEP_A |  -71.90678    23.3357    -3.08   0.002    -117.8522   -25.96138
       _cons |  -2630.703   5358.124    -0.49   0.624    -13180.25    7918.846
```

STATA.7

Source	SS	df	MS			
Model	390374743	3	130124914	Number of obs	=	272
Residual	4.7056e+09	268	17558366	F(3, 268)	=	7.41
Total	5.0960e+09	271	18804490.1	Prob > F	=	0.0001
				R-squared	=	0.0766
				Adj R-squared	=	0.0663
				Root MSE	=	4190.3

HICOSTR1_A	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AGEP_A	2.038146	35.71126	0.06	0.955	-68.27215	72.34844
EMPWRKLSWK_A	-956.8867	740.5141	-1.29	0.197	-2414.852	501.0784
medicare	-1752.913	884.9135	-1.98	0.049	-3495.179	-10.64606
_cons	5232.7	2022.65	2.59	0.010	1250.395	9215.005

STATA.8

Source	SS	df	MS			
Model	361056466	2	180528233	Number of obs	=	272
Residual	4.7350e+09	269	17602083.1	F(2, 269)	=	10.26
Total	5.0960e+09	271	18804490.1	Prob > F	=	0.0001
				R-squared	=	0.0709
				Adj R-squared	=	0.0639
				Root MSE	=	4195.5

HICOSTR1_A	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AGEP_A	-12.309	33.98362	-0.36	0.717	-79.21671	54.5987
medicare	-2131.864	835.9455	-2.55	0.011	-3777.692	-486.0356
_cons	5875.603	1962.944	2.99	0.003	2010.917	9740.29

STATA.9

Source	SS	df	MS			
Model	676485648	5	135297130	Number of obs	=	272
Residual	4.4195e+09	266	16614778.9	F(5, 266)	=	8.14
Total	5.0960e+09	271	18804490.1	Prob > F	=	0.0000
				R-squared	=	0.1327
				Adj R-squared	=	0.1164
				Root MSE	=	4076.1

HICOSTR1_A	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnFAMINCTC_A	1190.968	379.2261	3.14	0.002	444.3008	1937.634
WEIGHTLBTC_A	-17.80198	7.584576	-2.35	0.020	-32.73542	-2.868539
HISP_A	-2119.544	1104.821	-1.92	0.056	-4294.851	55.76334
AGEP_A	-41.19429	29.07597	-1.42	0.158	-98.44261	16.05403
EMPWRKLSWK_A	-1208.538	687.3202	-1.76	0.080	-2561.818	144.7421
_cons	-2840.355	5338.59	-0.53	0.595	-13351.62	7670.914

STATA.10

Source	SS	df	MS			
				Number of obs	=	177
				F(5, 171)	=	1.20
Model	59844843.2	5	11968968.6	Prob > F	=	0.3111
Residual	1.7050e+09	171	9970894.89	R-squared	=	0.0339
				Adj R-squared	=	0.0057
Total	1.7649e+09	176	10027658.3	Root MSE	=	3157.7

HICOSTR1_A	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnFAMINCTC_A	577.1158	353.7139	1.63	0.105	-121.0922	1275.324
WEIGHTLBTC_A	-6.24771	7.932037	-0.79	0.432	-21.90503	9.409607
HISP_A	-676.126	1325.414	-0.51	0.611	-3292.406	1940.154
AGEP_A	16.24089	41.91142	0.39	0.699	-66.4895	98.97127
EMPWRKLSWK_A	-853.7937	686.0436	-1.24	0.215	-2207.998	500.4111
_cons	-2961.437	5661.213	-0.52	0.602	-14136.3	8213.423

STATA.11

Source	SS	df	MS			
				Number of obs	=	95
				F(5, 89)	=	3.09
Model	426393340	5	85278668.1	Prob > F	=	0.0129
Residual	2.4588e+09	89	27627107.2	R-squared	=	0.1478
				Adj R-squared	=	0.0999
Total	2.8852e+09	94	30693679.6	Root MSE	=	5256.1

HICOSTR1_A	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnFAMINCTC_A	2511.767	911.6924	2.76	0.007	700.2531	4323.28
WEIGHTLBTC_A	-30.13009	14.89164	-2.02	0.046	-59.71946	-.5407178
HISP_A	-2434.291	1913.053	-1.27	0.207	-6235.486	1366.905
AGEP_A	43.67009	75.36412	0.58	0.564	-106.0768	193.417
EMPWRKLSWK_A	-518.8264	1547.81	-0.34	0.738	-3594.293	2556.64
_cons	-20029.53	12254.65	-1.63	0.106	-44379.27	4320.205